UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/645,210	08/21/2003	Xiaofan Lin	200300184-1	7233

22879 7590 05/25/2007

HEWLETT PACKARD COMPANY
P O BOX 272400, 3404 E. HARMONY ROAD
INTELLECTUAL PROPERTY ADMINISTRATION
FORT COLLINS, CO 80527-2400

EXAMINER	
ALBERTALLI, BRIAN LOUIS	
ART UNIT	**PAPER NUMBER**
2626	

MAIL DATE	DELIVERY MODE
05/25/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/645,210	LIN, XIAOFAN	
	Examiner	Art Unit	
	Brian L. Albertalli	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-19 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-4 and 6-19 is/are rejected.

7)☒ Claim(s) 5 is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 112

1. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

2. Claim 12 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 12 requires selecting a frame for extraction "if the frame's fitness datum exceeds a greatest fitness datum". This requirement is logically inconsistent. That is, within a set of fitness data the greatest fitness datum could not be exceeded, simply because that fitness datum is the greatest.

Paragraph 22 of the specification describes the selection process most likely intended to be described in claim 12. Here, it is made clear that a frame is selected when a given frame exceeds the greatest fitness datum minus a predetermined margin. Thus, when a fitness datum for a frame is above the greatest fitness value minus the predetermined margin (e.g. a greatest datum 12.0 minus a predetermined margin 2.0 would give a threshold of 10.0), then the frame is selected.

For the purposes of examination, claim 12 has been interpreted as described above in reference to paragraph 22 of the specification.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

4. Claims 1, 2, 8-12, and 19 are rejected under 35 U.S.C. 102(e) as being anticipated by Jiang et al. (U.S. Patent 6,901,362).

In regard to claims 1 and 19, Jiang et al. disclose a method and system for sound signal classification (Fig. 5), comprising steps/means for:

receiving a sound signal (step 302, column 12, lines 39-40);

specifying meta-data to be extracted from the sound signal (who is speaking, column 3, lines 38-44);

dividing the sound signal into a set of frames (step 304, column 12, lines 40-41);

applying a fitness function to the frames to create a set of fitness data (a "fitness function" is defined by the specification as "a mathematical calculation to be performed on one or more sound signal frames"; in step 308, various mathematical calculations are applied to the frames of speech, column 12, lines 41-44 and column 7, lines 15-18);

selecting a frame from the set of frames, if the frame's corresponding fitness datum within the set of fitness data exceeds a predetermined threshold value (frames are compared to a threshold to make a speech/non-speech determination, column 7,

lines 18-20; the frames that are classified as speech are selected to perform speaker change metadata from the speech frames, column 12, lines 45-60 and column 3, lines 39-41);

extracting the meta-data from the selected frames (the frames that are classified as speech are selected to perform speaker change metadata from the speech frames, column 12, lines 45-60 and column 3, lines 39-41); and

classifying the sound signal based on the meta-data extracted from the selected frames (identifying whether the speaker has changed, column 13, lines 10-11).

In regard to claim 2, Jiang et al. disclose the sound signal is a speech signal (determined to be speech, column 12, lines 45-60 and column 3, lines 39-41).

In regard to claim 8, Jiang et al. disclose specifying identity meta-data (a change of speaker, column 12, lines 45-60 and column 3, lines 39-41).

In regard to claims 9 and 10, Jiang et al. disclose dividing the sound signal into a set of equal length time frames (25 ms time frames, column 5, lines 54-62).

In regard to claim 11, Jiang et al. disclose calculating a signal strength of the sound signal frame (energy features of the frame are used for speech/non-speech determination, column 8, lines 23-25).

In regard to claim 12, Jiang et al. disclose selecting a frame for meta-data extraction, if the frame's fitness datum exceeds a greatest fitness datum within the set of fitness data by a predetermined margin (see interpretation of claim 12 in the rejection under 35 U.S.C. 112, 2nd paragraph, above; a high zero crossing rate ratio is used to determine speech/non-speech, column 7, lines 43-48).

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 3, 4, 6, and 7 are rejected under 35 U.S.C. 103(a) as being unpatentable over Jiang et al., in view of Kanevsky et al. (U.S. Patent 6,665,644).

Jiang et al. disclose classifying a sound signal into several classes (frames determined to be non-speech are classified into various categories, column 3, lines 29-44) and further discloses such classifications are beneficial in audio information retrieval (column 1, lines 31-34).

Jiang et al. do not disclose that the specifying includes age, gender, accent, or dialect meta-data.

Kanevsky et al. disclose a method for sound signal classification, wherein specifying meta-data to be extracted from a sound signal includes age, gender, accent, or dialect meta-data (column 2, lines 16-21).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Jiang et al. to specify age, gender, accent, or dialect meta-data to be extracted from the sound signal, because such additional data is valuable for information retrieval (data mining), as taught by Kanevsky et al. (column 1, lines 49-55).

7. Claims 13 and 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Jiang et al., in view of Pawlewski et al. (U.S. Patent 5,583,961).

Jiang et al. do not disclose extracting the meta-data from the selected frames using a MLP neural network having an input layer with nodes corresponding to the sound signal's Mel-Cepstral components.

Pawlewski et al. disclose a method for extracting meta-data (identity) from a sound signal, wherein the meta data is extracted from the selected frames using a MLP neural network having an input layer with nodes corresponding to the sound signal's Mel-Cepstral components (mel-frequency cepstral coefficients, MFCC's, are extracted from a sound signal and fed to a MLP neural network for comparison to determine who is speaking, column 4, lines 61-63 and column 9, lines 35-47).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Jiang et al. to extract the meta-data using a MLP neural network having an input layer with nodes corresponding to the sound signal's Mel-Cepstral components, because Mel-Cepstral components more closely approximate the human auditory system's response and MLP neural networks provide a trainable model for

classification that improves over time, thus the classification of the meta-data would be more accurate.

8. Claims 15 and 16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Jiang et al., in view of Kittler et al. (*On Combining Classifiers*).

Jiang et al. do not disclose assigning the sound signal to that meta-data class to which a largest number of the selected frames have been assigned, or

adding together each of the selected frame's confidence scores for each meta-data class; and

assigning the sound signal to that meta-data class having a highest total confidence score.

Kittler et al. disclose various methods for assigning a sound signal to a meta-class (who is speaking, page 230, section 4.3), comprising:

assigning the sound signal to that meta-data class to which a largest number of the selected frames have been assigned (page 229, section 3.4, combining a plurality of classifiers by majority vote), and

adding together each of the selected frame's confidence scores for each meta-data class; and

assigning the sound signal to that meta-data class having a highest total confidence score (page 228, section 2.2, summing together probability of a plurality of classifiers and assigning the sound to the highest score).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Jiang et al. to combine the results of the classification of each frame by assigning the sound signal to that meta-data class to which a largest number of the selected frames have been assigned or assigning the sound signal to that meta-data class having a highest total confidence score, because various classifier combination schemes outperform a single best classifier, as taught by Kittler et al. (page 226, 1st column, 2nd paragraph).

9. Claim 17 is rejected under 35 U.S.C. 103(a) as being unpatentable over Jiang et al., in view of Official Notice.

Jiang et al. do not disclose assigning the sound signal to that meta-data class having a statistically longest run-length.

Official Notice is taken it is well-known in the art to combine a plurality of classification results by assigning a class based on the statistically longest run-length.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Jiang et al. to assign the sound signal to that meta-data class having a statistically longest run-length, because the class to which the longest string of frames were assigned would be the most likely class for the sound signal.

10. Claim 18 is rejected under 35 U.S.C. 103(a) as being unpatentable over Jiang et al., in view of Pawlewski et al., and further in view of Kittler et al.

Jiang et al. disclose a method for sound signal classification (Fig. 5), comprising:

receiving a sound signal (step 302, column 12, lines 39-40);

specifying meta-data to be extracted from the sound signal (who is speaking, column 3, lines 38-44);

dividing the sound signal into a set of equal length time frames (25 ms time frames, column 5, lines 54-62);

applying a fitness function to the frames to create a set of fitness data (a "fitness function" is defined by the specification as "a mathematical calculation to be performed on one or more sound signal frames"; in step 308, various mathematical calculations are applied to the frames of speech, column 12, lines 41-44 and column 7, lines 15-18); and

selecting a frame for meta-data extraction, if the frame's fitness datum exceeds a greatest fitness datum within the set of fitness data by a predetermined margin (see interpretation of claim 12 in the rejection under 35 U.S.C. 112, 2nd paragraph, above; a high zero crossing rate ratio is used to determine speech/non-speech, column 7, lines 43-48).

Jiang et al. do not disclose the meta-data from the selected frames using a Multi-Layer Perceptron (MLP) neural network.

Pawlewski et al. disclose a method for extracting meta-data (identity) from a sound signal, wherein the meta data is extracted from the selected frames using a MLP neural network having an input layer with nodes corresponding to the sound signal's Mel-Cepstral components (mel-frequency cepstral coefficients, MFCC's, are extracted

from a sound signal and fed to a MLP neural network for comparison to determine who is speaking, column 4, lines 61-63 and column 9, lines 35-47).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Jiang et al. to extract the meta-data using a MLP neural network having an input layer with nodes corresponding to the sound signal's Mel-Cepstral components, because Mel-Cepstral components more closely approximate the human auditory system's response and MLP neural networks provide a trainable model for classification that improves over time, thus the classification of the meta-data would be more accurate.

Jiang et al. and Pawlewski et al. do not disclose adding together each of the selected frame's confidence scores for each meta-data class; and

assigning the sound signal to that meta-data class having a highest total confidence score.

Kittler et al. disclose adding together each of the selected frame's confidence scores for each meta-data class; and

assigning the sound signal to that meta-data class having a highest total confidence score (page 228, section 2.2, summing together probability of a plurality of classifiers and assigning the sound to the highest score).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Jiang et al. and Pawlewski et al. to combine the results of the classification of each frame by assigning the sound signal to that meta-data class having a highest total confidence score, because various classifier combination

schemes outperform a single best classifier, as taught by Kittler et al. (page 226, 1st column, 2nd paragraph).

Allowable Subject Matter

11. Claim 5 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The following is a statement of reasons for the indication of allowable subject matter: There is no teaching or suggestion in Jiang et al., Kanevsky et al., Pawlewski et al., or Kittler et al. to set the threshold for selecting a set of frames for classifying to a ratio of between about 1:2 and 1:3. While Jiang et al. disclose selecting a subset of frames for classification, there is no indication that a specific ratio of frames is selected from the sound signal. Rather, the selection is based on whether speech is present or not. Pawlewski et al., Kanevsky et al., and Kittler et al. provide no teaching of selecting a subset of frames for classification.

Conclusion

12. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Wang (U.S. Patent 5,596,679) disclose a method for classifying sounds by a voting scheme. Kanevsky et al. (U.S. Patent 6,442,519) discloses a

system that utilizes extracted meta-data to update speech models. Beattie et al. (U.S. Patent 5,865,626) disclose a method for identifying dialect meta-data. Yamamoto (U.S. Patent 6,122,615) discloses a speech recognizer that utilizes gender meta-data in a re-recognition process. Fussell (*Automatic Sex Identification From Short Segments of Speech*) discloses a method for extracting gender meta-data from frames of speech. Chan et al. (*Classification of Speech Accents with Neural Networks*) disclose a method for identifying accent meta-data using neural networks.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L. Albertalli whose telephone number is (571) 272-7616. The examiner can normally be reached on Mon - Fri, 8:00 AM - 5:30 PM, every second Fri off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

BLA 5/22/07

DAVID HUDSPETH
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER